

DOCUMENT RESUME

ED 264 275

TM 850 761

AUTHOR Sympson, J. B.; Hartmann, Lorelee
TITLE Item Calibrations for Computerized Adaptive Testing (CAT) Experimental Item Pools.
INSTITUTION Naval Personnel Research and Development Lab., Washington, D.C.
PUB DATE Jul 82
NOTE 4p.; In: Item Response Theory and Computerized Adaptive Testing Conference Proceedings (Wayzata, MN, July 27-30, 1982) (TM 850 744).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Adaptive Testing; Adults; *Computer Assisted Testing; Estimation (Mathematics); *Item Analysis; *Item Banks; Latent Trait Theory; Psychometrics; Research Methodology; *Test Construction
IDENTIFIERS *Armed Services Vocational Aptitude Battery; LOGIST Computer Program; *Parametric Analysis

ABSTRACT

In connection with the development of item pools for the Navy's experimental Computerized Adaptive Testing (CAT) system, items in five different content areas (General Science, Paragraph Composition, Word Knowledge, Math Knowledge, and Arithmetic Reasoning) have been calibrated using item response theory (IRT) methodology. Operational Armed Services Vocational Aptitude Battery (ASVAB) subtests 8, 9, and 10 were calibrated with the new CAT items to insure that all estimated parameters in a content area were expressed relative to a common metric. The computer program LOGIST was used to fit the 3-parameter logistic model to the available item response data. Modifications to version 2.B of this program were made prior to completing the calibrations. These modifications were tested and found to improve the parameter estimates generated by LOGIST. In each content area, all CAT items and ASVAB items were calibrated simultaneously in a single LOGIST run. The median values of the estimated IRT parameters, obtained in each content area, suggest that CAT procedures will function quite effectively in each of these five content areas. Items in four more content areas, as well as additional items will be calibrated to augment three of the previously developed item banks. (PN)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ITEM CALIBRATIONS FOR
COMPUTERIZED ADAPTIVE TESTING (CAT)
EXPERIMENTAL ITEM POOLS

J. B. SYMPSON AND LORALEE HARTMANN
NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☐ This document has been reproduced as
received from the person or organization
originating it.

☒ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

In connection with the development of item pools for the Navy's experimen-
tal CAT system, items in five different content areas (General Science, Para-
graph Comprehension, Word Knowledge, Math Knowledge, and Arithmetic Reasoning)
have been calibrated using item response theory (IRT) methodology. Calibration
of these items was originally undertaken by a government contractor (Research
Applications, Inc.), but a review of the procedures followed by the contractor
indicated that the calibrations should be repeated.

In the calibrations conducted at Navy Personnel Research and Development
Center (NPRDC), operational Armed Services Vocational Aptitude Battery (ASVAB)
subtests 8, 9, and 10 were calibrated along with the new computerized adaptive
testing (CAT) items in order to insure that all estimated parameters in a con-
tent area were expressed relative to a common metric.

An example of the "linking design" for one content area (Arithmetic Reason-
ing) is shown in Table 1. Each row of this table (i.e., each "Group") repre-
sents approximately 312 male applicants for military service who completed one
particular experimental Arithmetic Reasoning (AR) booklet and one particular
operational ASVAB form. The experimental booklet and the ASVAB form completed
by individuals in a given row of the table are indicated by Xs. Each experimen-
tal AR booklet contained 35 items and each ASVAB form contained 30 AR items.
Thus, a total of 390 AR items are represented in the table. Each individual in
the calibration sample was exposed to 65 of these items. An average of approxi-
mately 1,870 responses were available for each of the 390 AR items.

The number of experimental booklets, the number of CAT items, the number of
ASVAB items, the total number of items, and the number of individuals in the
final calibration sample for each content area are given in Table 2. As indi-
cated by the column totals, the responses of 34,774 individuals were used to
obtain IRT parameters for a total of 1,750 test items.

The computer program LOGIST, developed by Lord at Educational Testing Ser-
vice, was used to fit the 3-parameter logistic model to the available item re-
sponse data. Version 2.8 of this program (released in 1976) was used, but sev-
eral modifications to the program were made prior to completing the calibra-
tions. These modifications were tested at NPRDC and found to improve the param-

Table 1
Linking Design for Arithmetic Reasoning Items

Group	CAT Booklet						ASVAB Form					
	B1	B2	B3	B4	B5	B6	8A	8B	9A	9B	10A	10B
1	X						X					
2	X							X				
3	X								X			
4	X									X		
5	X										X	
6	X											X
7		X					X					
8		X						X				
9		X							X			
10		X								X		
11		X									X	
12		X										X
13			X				X					
14			X					X				
15			X						X			
16			X							X		
17			X								X	
18			X									X
19				X			X					
20				X				X				
21				X					X			
22				X						X		
23				X							X	
24				X								X
25					X		X					
26					X			X				
27					X				X			
28					X					X		
29					X						X	
30					X							X
31						X	X					
32						X		X				
33						X			X			
34						X				X		
35						X					X	
36						X						X

eter estimates generated by LOGIST. In each content area, all CAT items and ASVAB items were calibrated simultaneously in a single LOGIST run.

Table 2
Characteristics of Calibration Database

Content Area	No. of CAT Booklets	CAT Items	ASVAB Items	Total Items	Sample Size
GS	2	200	150	350	4106
PC	4	180	90	270	7734
WK	2	200	210	410	4225
MK	4	180	150	330	7491
AR	6	210	180	390	11218
Total	18	970	780	1750	34774

Table 3 shows median values of the estimated IRT parameters $\hat{a}$, $\hat{b}$, and $\hat{c}$ that were obtained in each content area, separately for CAT items and operational ASVAB items. The median $\hat{a}$ values in column 1 of this table are high enough, and the median $\hat{c}$ values in column 3 are low enough, to suggest that CAT procedures will function quite effectively in each of these five content areas.

Table 3
Median Values of Estimated IRT Parameters

Content Area	CAT Items			ASVAB Items		
	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{a}$	$\hat{b}$	$\hat{c}$
GS	1.29	.02	.22	1.05	-.05	.22
PC	1.22	-.06	.23	.87	-.56	.23
WK	1.41	-.74	.25	1.45	-.15	.25
MK	1.38	.66	.18	1.66	.69	.19
AR	1.39	-.01	.22	1.52	.28	.22

During the next few months, items in four more content areas (Automotive Information, Electronics Information, Mechanical Comprehension, and Shop Information) will be calibrated using the procedures described above. Additional items will also be calibrated in order to augment three of the previously developed item banks.